DDocE: Deep Document Enhancement with Multi-Scale Feature Aggregation and Pixel-Wise Adjustments

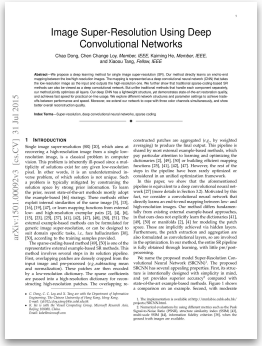
Karina O. M. Bogdan, Guilherme A. S. Megeto, Rovilson Leal, Gustavo Souza, Augusto C. Valente, and Lucas N. Kirsten

CBDAR 2021

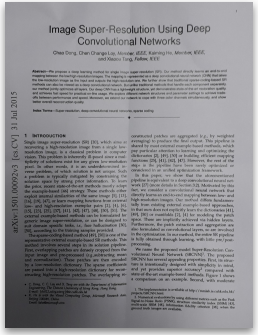
ELDORADO

hp

Document Enhancement

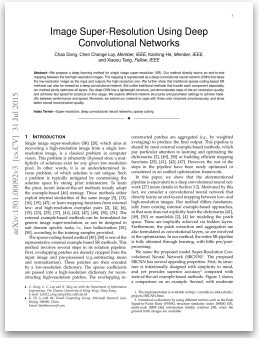


PDF



Photo

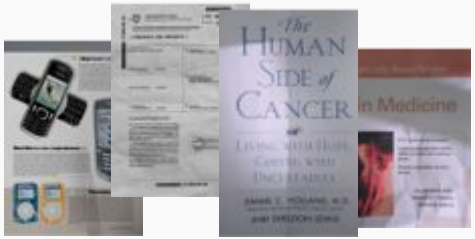
Document Scanner Model



Enhanced Image

What do we want to enhance?

Uneven illumination



Paper-related artifacts

Goal: enhance and maintain the text readability

Image Enhancement

Current State-of-the-Art

Zero-DCE[1]

CURL[2]

MIRNet[3]

Focus on natural images

Input GT BEDSR-Net[4]

Focus on shadows only

[1] Guo et al. Zero-Reference Deep Curve Estimation for Low-Light Image Enhancement. In CVPR, 2020.
[2] Moran et al. CURL: Neural Curve Layers for Global Image Enhancement. In ICPR, 2020.
[3] Zamir et al. Learning Enriched Features for Real Image Restoration and Enhancement. In ECCV, 202.
[4] Lin et al. BEDSR-Net: A Deep Shadow Removal Network from a Single Document Image. In CVPR, 2020.

Document Enhancement

Controlled environments

DeepOtsu[1]
Focus on binarization

Robust models

Fan[2]
Watershed-based segmentation

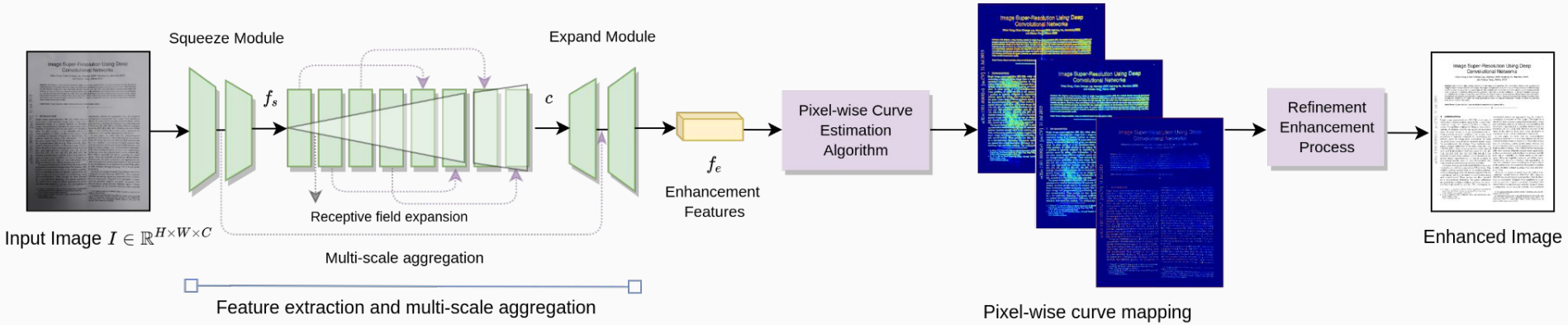
HP Inc.[3]
Deep learning ensemble model with segmentation maps

[1] He and Schomaker. DeepOtsu: Document Enhancement and Binarization using Iterative Deep Learning. In Pattern Recog, 2019.
[2] Fan. Enhancement of Camera-captured Document Images with Watershed Segmentation. In CBDAR, 2007.
[3] HP Inc.. A workflow for document enhancement through content segmentation and multiple enhancements. Technical Disclosure Commons, 2020.

DDocE: Deep Document Enhancement

Context and Multi-Scale Feature Aggregator

Pixel-Wise Adjustments

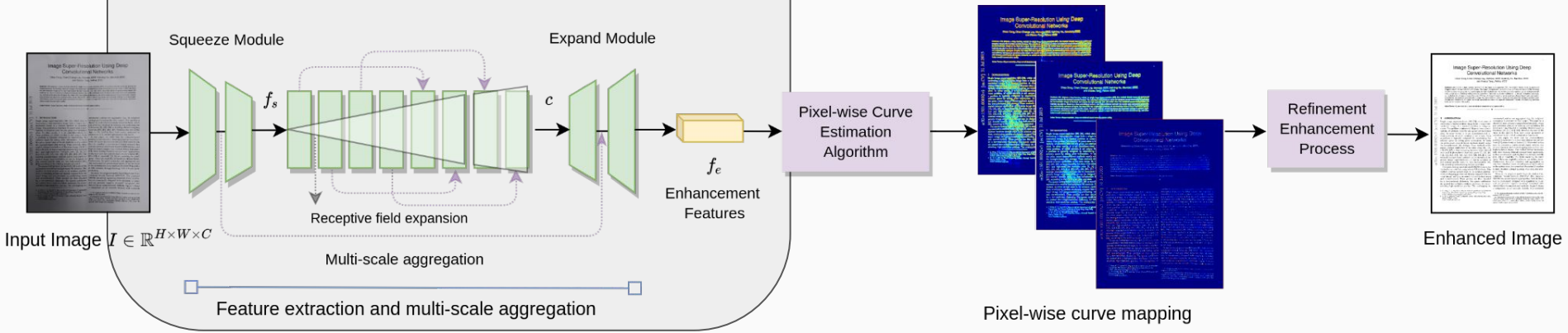


DDocE: Deep Document Enhancement

Context and Multi-Scale Feature Aggregator

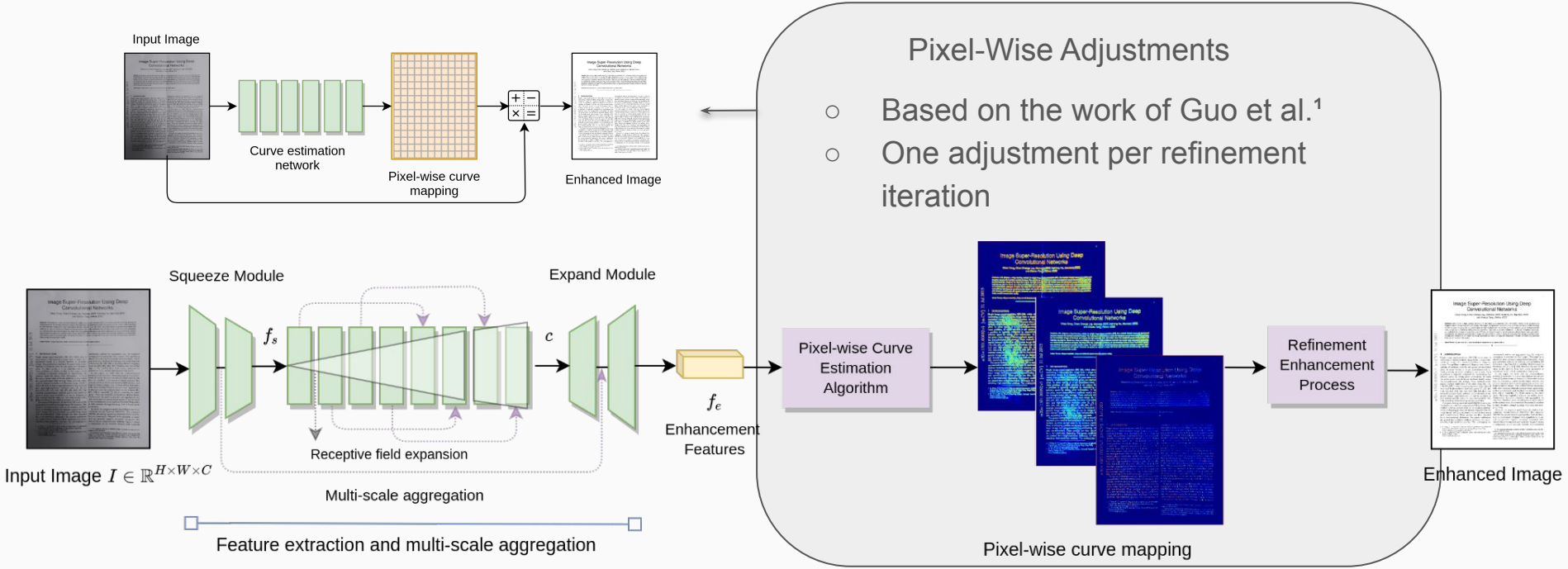
- Adapted from Yu et al.[1]
- Receptive field expansion with dilated convolutions

Layer	1	2	3	4	5	6	7	8	Output
Convolution	3×3	3×3	3×3	3×3	3×3	3×3	3×3	3×3	-
Dilation factor	1	1	2	3	8	16	1	1	-
Receptive field	3×3	5×5	9×9	17×17	33×33	65×65	67×67	69×69	-
Output channels	C_s	C_s	C_s	C_s	C_s	C_s	C_s	C_s	$2C_s$
Input channels	C_s	C_s	C_s	C_s	C_s	$2C_s$	$2C_s$	$2C_s$	$2C_s$
Skip connection (concatenated layer)	-	-	-	-	1	2	3	4	-



[1] Yu et al. Multi-Scale Context Aggregation by Dilated Convolutions. In International Conference on Learning Representations (ICLR), 2016.
[2] Guo et al. Zero-Reference Deep Curve Estimation for Low-Light Image Enhancement. In CVPR, 2020.

DDocE: Deep Document Enhancement



[1] Guo et al. Zero-Reference Deep Curve Estimation for Low-Light Image Enhancement. In CVPR, 2020.

Experiments

Dataset

- ✚ Document variation
- ➖ Small dataset: ~770 pairs
  - “Fake” ground-truth

Data augmentation

- ✚ Cropping
- ✚ Pyramid Sliding Window

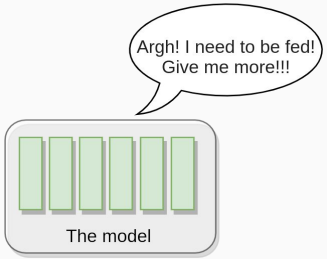


Image Quality Assessment (IQA)

- Traditional
  - PSNR
  - Multi-Scale SSIM
- Deep learning-based
  - PieAPP[1]
  - WaDIQaM[2]
  - LPIPS[3]
  - DISTS[4]

[1] Prashnani et al., PieAPP: Perceptual Image-Error Assessment through Pairwise Preference. In CVPR, 2018.
[2] Bosse et al., Deep Neural Networks for No-Reference and Full-Reference Image Quality Assessment. In IEEE Transactions on Image Processing, 2018.
[3] Zhang et al., The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. In CVPR, 2018
[4] Ding et al. Image Quality Assessment: Unifying Structure and Texture Similarity. In arXiv:2004.07728, 2020

Quantitative Results

Model	#	PSNR ↑	MS-SSIM ↑	PieAPP↓	WaDIQaM ↑	LPIPS↓	DISTS↓
U-Net[1]	2M	15.9 +-2.7	0.81 +-.07	**0.89 +-0.7**	0.56 +-.10	0.17 +-.05	0.20 +-.05
Pix2Pix[2]	57M	16.2 +-2.7	0.81 +-.08	1.18 +-0.8	0.55 +-.13	0.17 +-.05	0.19 +-.05
Pix2PixHD[3]	113M	15.7 +-2.7	0.79 +-.09	0.95 +-1.1	0.51 +-.12	0.21 +-.06	0.24 +-.06
Fan[4]	-	15.1 +-3.4	0.80 +-.08	1.00 +-0.9	0.58 +-.08	0.16 +-.05	0.16 +-.05
HP Inc.[5]	5M	**16.9 +-3.0**	**0.84 +-.07**	1.49 +-0.9	0.57 +-.10	0.16 +-.05	0.17 +-.05
Zero-DCE[6]	79k	15.9 +-3.0	0.82 +-.08	1.10 +-0.8	0.58 +-.09	0.17 +-.05	0.19 +-.04
DDocE	595k	16.1 +-2.7	0.82 +-.07	1.09 +-0.9	0.58 +-.09	0.16 +-.05	0.17 +-.05
DDocE w/ pyramid	595k	16.1 +-3.0	0.83 +-.07	1.08 +-0.8	**0.63 +-.06**	**0.14 +-.05**	**0.15 +-.04**

Competitive models from other tasks

Enhancement Models

Our model

Best result is highlighted in bold and "#" represents the number of parameters in the model.

[1] Ronneberger et al. U-Net: convolutional networks for biomedical image segmentation. In: MICCAI, 2015.
[2] Isola et a.. Image-to-image translation with conditional adversarial networks. In: CVPR, 2017.
[3] Wang et al. High-resolution image synthesis and semantic manipulation with conditional gans. In: CVPR, 2018.
[4] Fan. Enhancement of Camera-captured Document Images with Watershed Segmentation. In CBDAR, 2007.
[5] HP Inc.. A workflow for document enhancement through content segmentation and multiple enhancements. Technical Disclosure Commons, 2020.
[6] Guo et al. Zero-Reference Deep Curve Estimation for Low-Light Image Enhancement. In CVPR, 2020.

Qualitative Results

Reconstruction-based

- ✚ Crumpled paper
- ▬ Documents with colored regions
- ▬ Artifacts on text regions
  Shadow artifacts

Enhancement Models

Fan[1]

- ✚ Crumpled paper
- ▬ Documents with colored regions

HP Inc.[2]

- ✚ Crumpled paper
- ▬ Documents with colored regions

Zero-DCE[3]

- ✚ Lightweight model
- ▬ Documents with colored regions
- ▬ Crumpled paper

DDoCE (Ours)

- ✚ Lightweight model
- ✚ Documents with colored regions
- ▬ Crumpled paper

[1] Fan. Enhancement of Camera-captured Document Images with Watershed Segmentation. In CBDAR, 2007.
[2] HP Inc.. A workflow for document enhancement through content segmentation and multiple enhancements. Technical Disclosure Commons, 2020.
[3] Guo et al. Zero-Reference Deep Curve Estimation for Low-Light Image Enhancement. In CVPR, 2020.

Qualitative Results

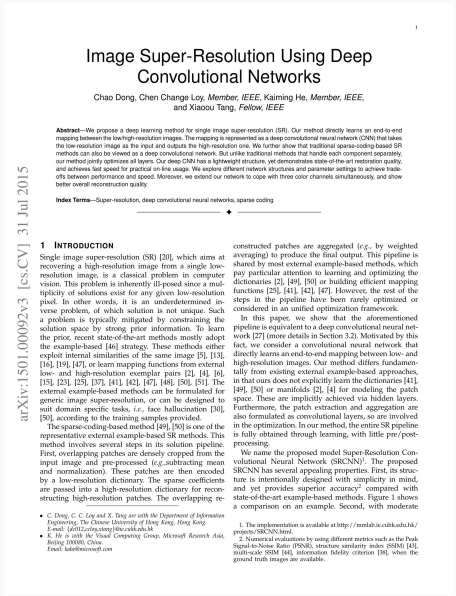


Image Super-Resolution Using Deep Convolutional Networks

Chao Dong, Chen Change Loy, *Member, IEEE*, Kaiming He, *Member, IEEE*, and Xiaoou Tang, *Fellow, IEEE*

Abstract—We propose a deep learning method for single image super-resolution (SR). Our method directly learns an end-to-end mapping between the low/high-resolution images. The mapping is represented as a deep convolutional neural network (CNN) that takes the low-resolution image as the input and outputs the high-resolution one. We further show that traditional sparse-coding-based SR methods can also be viewed as a deep convolutional network. But unlike traditional methods that handle each component separately, our method jointly optimizes all layers. Our deep CNN has a lightweight structure, yet demonstrates state-of-the-art restoration quality, and achieves fast speed for practical on-line usage. We explore different network structures and parameter settings to achieve trade-offs between performance and speed. Moreover, we extend our network to cope with three color channels simultaneously, and show better overall reconstruction quality.

Index Terms—Super-resolution, deep convolutional neural networks, sparse coding

1 Introduction

Single image super-resolution (SR) [20], which aims at recovering a high-resolution image from a single low-resolution image, is a classical problem in computer vision. This problem is inherently ill-posed since a multiplicity of solutions exist for any given low-resolution pixel. In other words, it is an underdetermined inverse problem, of which solution is not unique. Such a problem is typically mitigated by constraining the solution space by strong prior information. To learn the prior, recent state-of-the-art methods mostly adopt the example-based [46] strategy. These methods either exploit internal similarities of the same image [5], [13], [16], [19], [47], or learn mapping functions from external low- and high-resolution exemplar pairs [2], [4], [6], [15], [23], [25], [37], [41], [42], [47], [48], [50], [51]. The external example-based methods can be formulated for generic image super-resolution, or can be designed to suit domain specific tasks, *i.e.*, face hallucination [30], [50], according to the training samples provided.

The sparse-coding-based method [49], [50] is one of the representative external example-based SR methods. This method involves several steps in its solution pipeline. First, overlapping patches are densely cropped from the input image and pre-processed (*e.g.*,subtracting mean and normalization). These patches are then encoded by a low-resolution dictionary. The sparse coefficients are passed into a high-resolution dictionary for reconstructing high-resolution patches. The overlapping reconstructed patches are aggregated (*e.g.*, by weighted averaging) to produce the final output. This pipeline is shared by most external example-based methods, which pay particular attention to learning and optimizing the dictionaries [2], [49], [50] or building efficient mapping functions [25], [41], [42], [47]. However, the rest of the steps in the pipeline have been rarely optimized or considered in an unified optimization framework.

In this paper, we show that the aforementioned pipeline is equivalent to a deep convolutional neural network [27] (more details in Section 3.2). Motivated by this fact, we consider a convolutional neural network that directly learns an end-to-end mapping between low- and high-resolution images. Our method differs fundamentally from existing external example-based approaches, in that ours does not explicitly learn the dictionaries [41], [49], [50] or manifolds [2], [4] for modeling the patch space. These are implicitly achieved via hidden layers. Furthermore, the patch extraction and aggregation are also formulated as convolutional layers, so are involved in the optimization. In our method, the entire SR pipeline is fully obtained through learning, with little pre/post-processing.

We name the proposed model Super-Resolution Convolutional Neural Network (SRCNN)[1]. The proposed SRCNN has several appealing properties. First, its structure is intentionally designed with simplicity in mind, and yet provides superior accuracy[2] compared with state-of-the-art example-based methods. Figure 1 shows a comparison on an example. Second, with moderate

• C. Dong, C. C. Loy and X. Tang are with the Department of Information Engineering, The Chinese University of Hong Kong, Hong Kong.
E-mail: {dc012,ccloy,xtang}@ie.cuhk.edu.hk
• K. He is with the Visual Computing Group, Microsoft Research Asia, Beijing 100080, China.
Email: kahe@microsoft.com

1. The implementation is available at http://mmlab.ie.cuhk.edu.hk/projects/SRCNN.html.
2. Numerical evaluations by using different metrics such as the Peak Signal-to-Noise Ratio (PSNR), structure similarity index (SSIM) [43], multi-scale SSIM [44], information fidelity criterion [38], when the ground truth images are available.

PDF

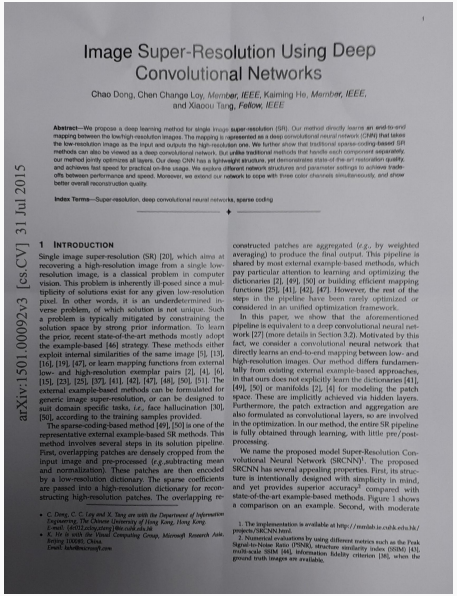


Image Super-Resolution Using Deep Convolutional Networks

Chao Dong, Chen Change Loy, *Member, IEEE*, Kaiming He, *Member, IEEE*, and Xiaoou Tang, *Fellow, IEEE*

Abstract—We propose a deep learning method for single image super-resolution (SR). Our method directly learns an end-to-end mapping between the low/high-resolution images. The mapping is represented as a deep convolutional neural network (CNN) that takes the low-resolution image as the input and outputs the high-resolution one. We further show that traditional sparse-coding-based SR methods can also be viewed as a deep convolutional network. But unlike traditional methods that handle each component separately, our method jointly optimizes all layers. Our deep CNN has a lightweight structure, yet demonstrates state-of-the-art restoration quality, and achieves fast speed for practical on-line usage. We explore different network structures and parameter settings to achieve trade-offs between performance and speed. Moreover, we extend our network to cope with three color channels simultaneously, and show better overall reconstruction quality.

Index Terms—Super-resolution, deep convolutional neural networks, sparse coding

1 Introduction

Single image super-resolution (SR) [20], which aims at recovering a high-resolution image from a single low-resolution image, is a classical problem in computer vision. This problem is inherently ill-posed since a multiplicity of solutions exist for any given low-resolution pixel. In other words, it is an underdetermined inverse problem, of which solution is not unique. Such a problem is typically mitigated by constraining the solution space by strong prior information. To learn the prior, recent state-of-the-art methods mostly adopt the example-based [46] strategy. These methods either exploit internal similarities of the same image [5], [13], [16], [19], [47], or learn mapping functions from external low- and high-resolution exemplar pairs [2], [4], [6], [15], [23], [25], [37], [41], [42], [47], [48], [50], [51]. The external example-based methods can be formulated for generic image super-resolution, or can be designed to suit domain specific tasks, *i.e.*, face hallucination [30], [50], according to the training samples provided.

The sparse-coding-based method [49], [50] is one of the representative external example-based SR methods. This method involves several steps in its solution pipeline. First, overlapping patches are densely cropped from the input image and pre-processed (*e.g.*,subtracting mean and normalization). These patches are then encoded by a low-resolution dictionary. The sparse coefficients are passed into a high-resolution dictionary for reconstructing high-resolution patches. The overlapping reconstructed patches are aggregated (*e.g.*, by weighted averaging) to produce the final output. This pipeline is shared by most external example-based methods, which pay particular attention to learning and optimizing the dictionaries [2], [49], [50] or building efficient mapping functions [25], [41], [42], [47]. However, the rest of the steps in the pipeline have been rarely optimized or considered in an unified optimization framework.

In this paper, we show that the aforementioned pipeline is equivalent to a deep convolutional neural network [27] (more details in Section 3.2). Motivated by this fact, we consider a convolutional neural network that directly learns an end-to-end mapping between low- and high-resolution images. Our method differs fundamentally from existing external example-based approaches, in that ours does not explicitly learn the dictionaries [41], [49], [50] or manifolds [2], [4] for modeling the patch space. These are implicitly achieved via hidden layers. Furthermore, the patch extraction and aggregation are also formulated as convolutional layers, so are involved in the optimization. In our method, the entire SR pipeline is fully obtained through learning, with little pre/post-processing.

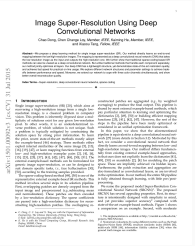
We name the proposed model Super-Resolution Convolutional Neural Network (SRCNN)[1]. The proposed SRCNN has several appealing properties. First, its structure is intentionally designed with simplicity in mind, and yet provides superior accuracy[2] compared with state-of-the-art example-based methods. Figure 1 shows a comparison on an example. Second, with moderate

• C. Dong, C. C. Loy and X. Tang are with the Department of Information Engineering, The Chinese University of Hong Kong, Hong Kong.
E-mail: {dc012,ccloy,xtang}@ie.cuhk.edu.hk
• K. He is with the Visual Computing Group, Microsoft Research Asia, Beijing 100080, China.
Email: kahe@microsoft.com

1. The implementation is available at http://mmlab.ie.cuhk.edu.hk/projects/SRCNN.html.
2. Numerical evaluations by using different metrics such as the Peak Signal-to-Noise Ratio (PSNR), structure similarity index (SSIM) [43], multi-scale SSIM [44], information fidelity criterion [38], when the ground truth images are available.

Photo (Input)

Qualitative Results



PDF

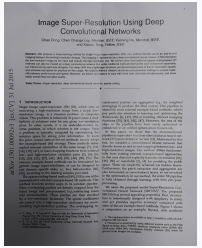
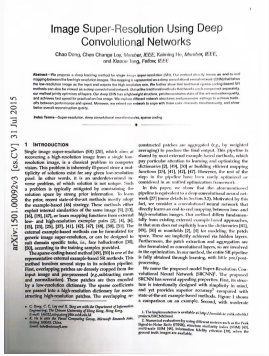
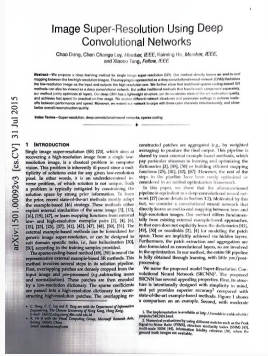


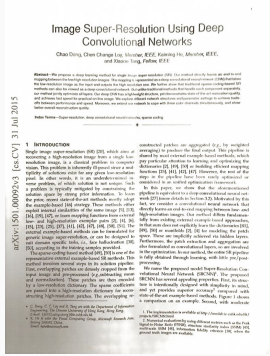
Photo Input



U-Net

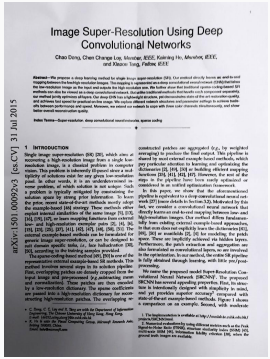


Pix2Pix

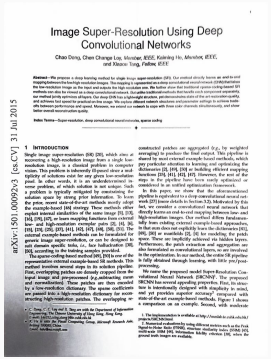


Pix2PixHD

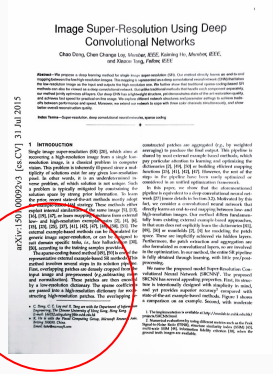
HP Inc.



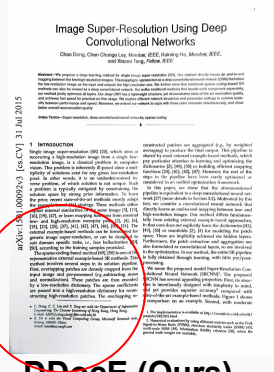
Fan



Zero-DCE

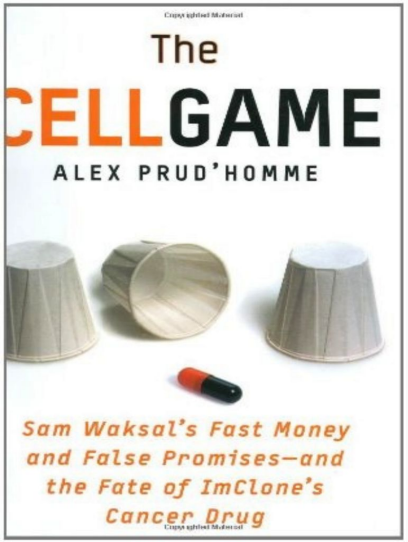


DDocE (Ours)



DDocE (Ours) w/ pyramid

Qualitative Results



PDF

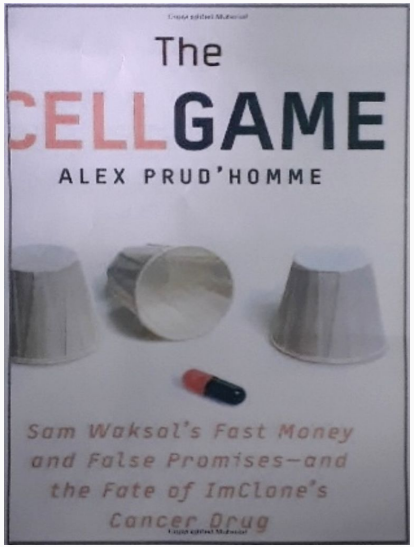
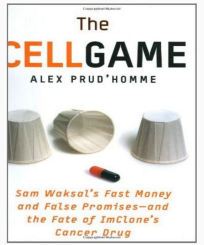


Photo (Input)

Qualitative Results



PDF

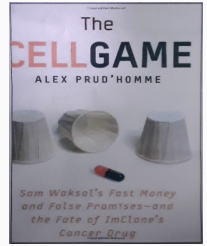
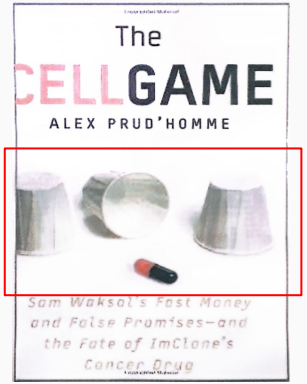
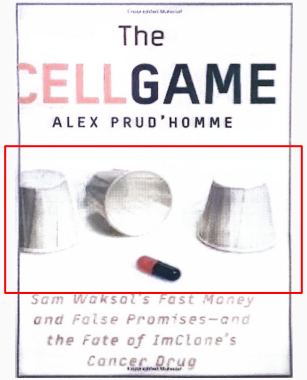


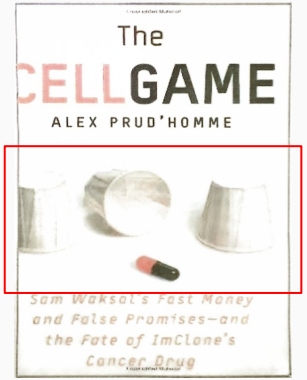
Photo Input



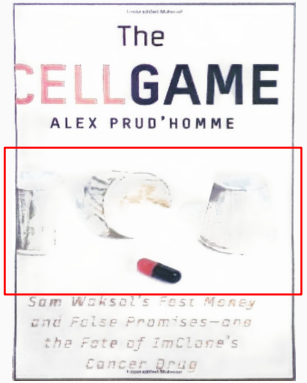
U-Net



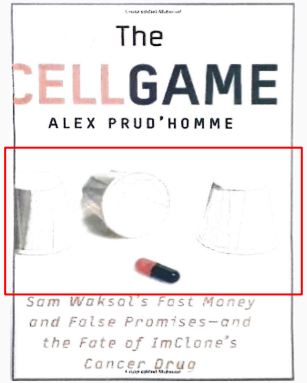
Pix2Pix



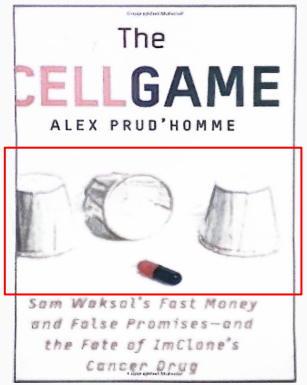
Pix2PixHD



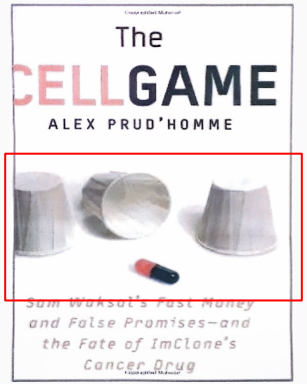
HP Inc.



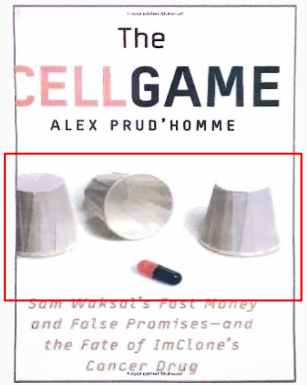
Fan



Zero-DCE



DDocE (Ours)



DDocE (Ours) w/ pyramid

Qualitative Results

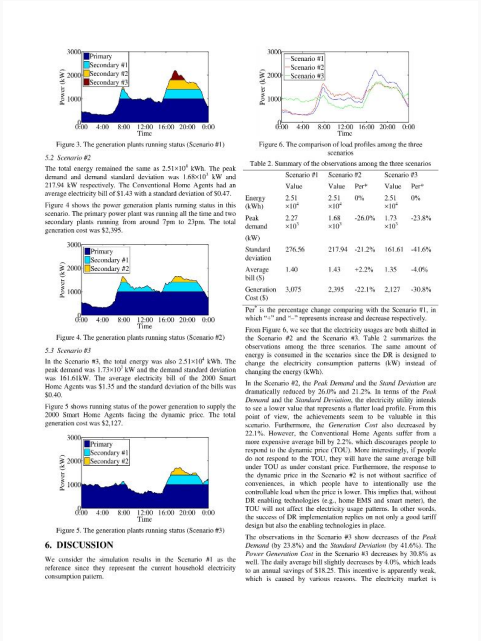


Figure 3. The generation plants running status (Scenario #1)

5.2 Scenario #2

The total energy remained the same as 2.51×10^5 kWh. The peak demand and demand standard deviation was 1.68×10^3 kW and 217.94 kW respectively. The Conventional Home Agents had an average electricity bill of $1.43 with a standard deviation of $0.47.

Figure 4 shows the power generation plants running status in this scenario. The primary power plant was running all the time and two secondary plants running from around 7pm to 23pm. The total generation cost was $2,395.

Figure 4. The generation plants running status (Scenario #2)

5.3 Scenario #3

In the Scenario #3, the total energy was also 2.51×10^5 kWh. The peak demand was 1.73×10^3 kW and the demand standard deviation was 161.61kW. The average electricity bill of the 2000 Smart Home Agents was $1.35 and the standard deviation of the bills was $0.40.

Figure 5 shows running status of the power generation to supply the 2000 Smart Home Agents facing the dynamic price. The total generation cost was $2,127.

Figure 5. The generation plants running status (Scenario #3)

6. DISCUSSION

We consider the simulation results in the Scenario #1 as the reference since they represent the current household electricity consumption pattern.

Figure 6. The comparison of load profiles among the three scenarios

Table 2. Summary of the observations among the three scenarios

	Scenario #1	Scenario #2		Scenario #3	
	Value	Value	Per*	Value	Per*
Energy (kWh)	2.51 $\times10^5$	2.51 $\times10^5$	0%	2.51 $\times10^5$	0%
Peak demand (kW)	2.27 $\times10^3$	1.68 $\times10^3$	-26.0%	1.73 $\times10^3$	-23.8%
Standard deviation	276.56	217.94	-21.2%	161.61	-41.6%
Average bill ($)	1.40	1.43	+2.2%	1.35	-4.0%
Generation Cost ($)	3,075	2,395	-22.1%	2,127	-30.8%

Per* is the percentage change comparing with the Scenario #1, in which "+" and "-" represents increase and decrease respectively.

From Figure 6, we see that the electricity usages are both shifted in the Scenario #2 and the Scenario #3. Table 2 summarizes the observations among the three scenarios. The same amount of energy is consumed in the scenarios since the DR is designed to change the electricity consumption patterns (kW) instead of changing the energy (kWh).

In the Scenario #2, the *Peak Demand* and the *Stand Deviation* are dramatically reduced by 26.0% and 21.2%. In terms of the *Peak Demand* and the *Standard Deviation*, the electricity utility intends to see a lower value that represents a flatter load profile. From this point of view, the achievements seem to be valuable in this scenario. Furthermore, the *Generation Cost* also decreased by 22.1%. However, the Conventional Home Agents suffer from a more expensive average bill by 2.2%, which discourages people to respond to the dynamic price (TOU). More interestingly, if people do not respond to the TOU, they will have the same average bill under TOU as under constant price. Furthermore, the response to the dynamic price in the Scenario #2 is not without sacrifice of conveniences, in which people have to intentionally use the controllable load when the price is lower. This implies that, without DR enabling technologies (e.g., home EMS and smart meter), the TOU will not affect the electricity usage patterns. In other words, the success of DR implementation replies on not only a good tariff design but also the enabling technologies in place.

The observations in the Scenario #3 show decreases of the *Peak Demand* (by 23.8%) and the *Standard Deviation* (by 41.6%). The *Power Generation Cost* in the Scenario #3 decreases by 30.8% as well. The daily average bill slightly decreases by 4.0%, which leads to an annual savings of $18.25. This incentive is apparently weak, which is caused by various reasons. The electricity market is

PDF

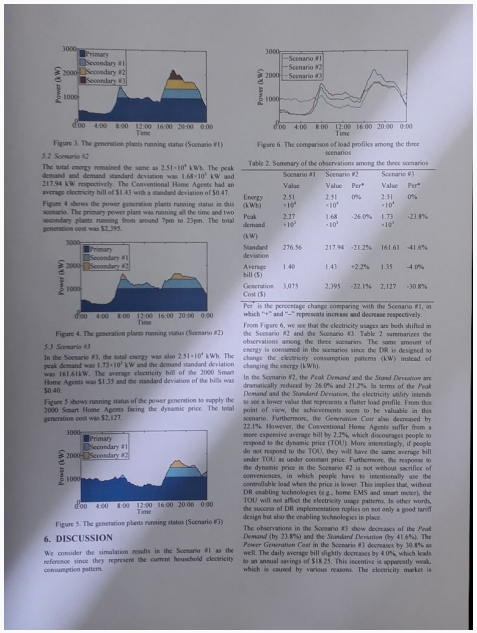
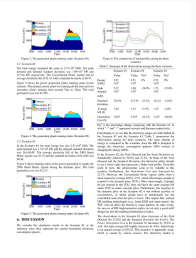


Photo (Input)

Qualitative Results



PDF

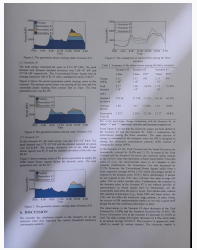
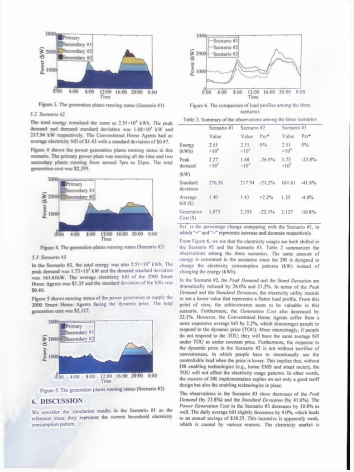
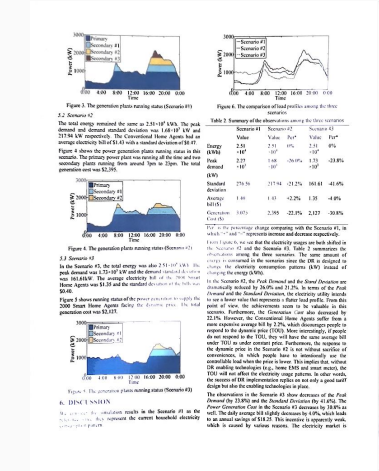


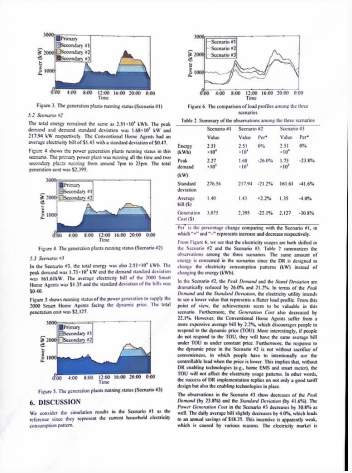
Photo Input



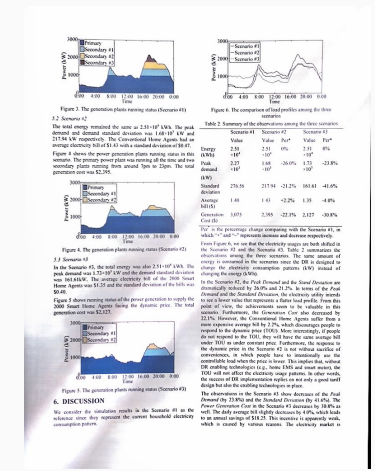
HP Inc.



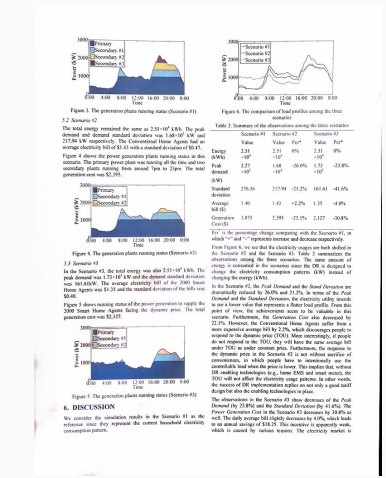
Fan



Zero-DCE

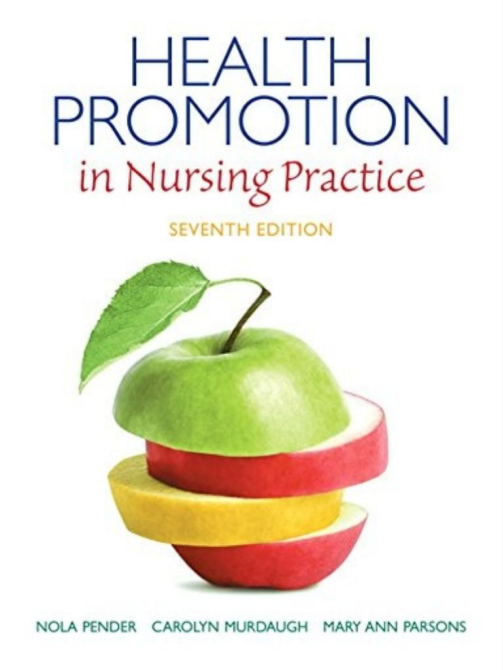


DDocE (Ours)



DDocE (Ours) w/ pyramid

Qualitative Results



PDF

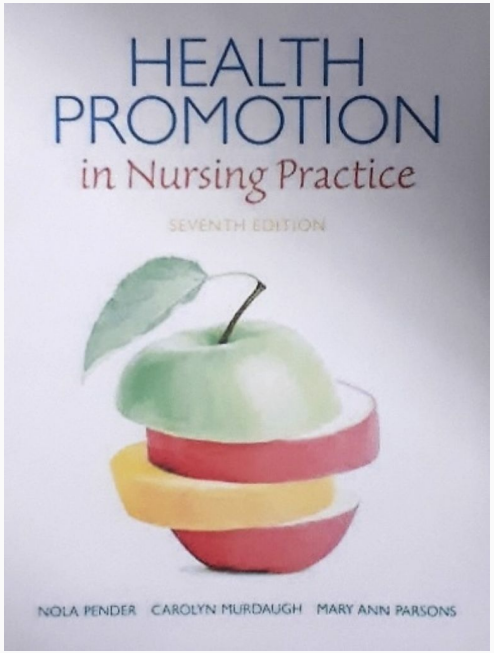
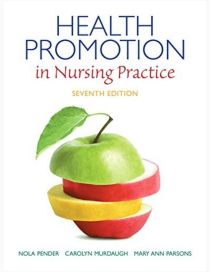


Photo (Input)

Qualitative Results



PDF

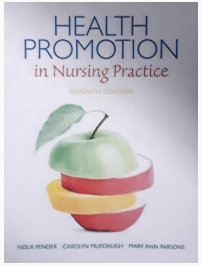
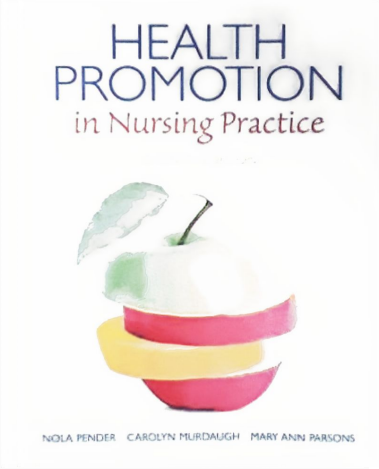
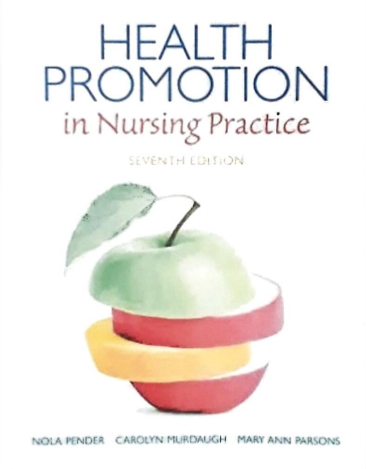


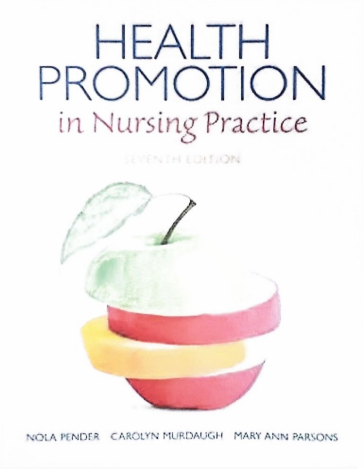
Photo Input



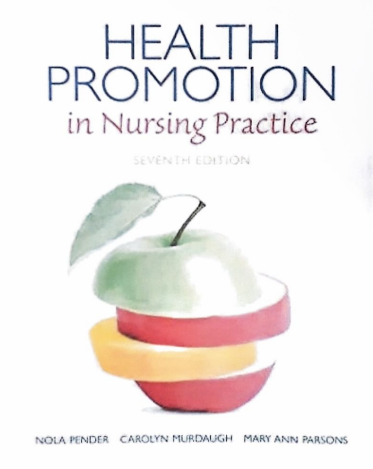
HP Inc.



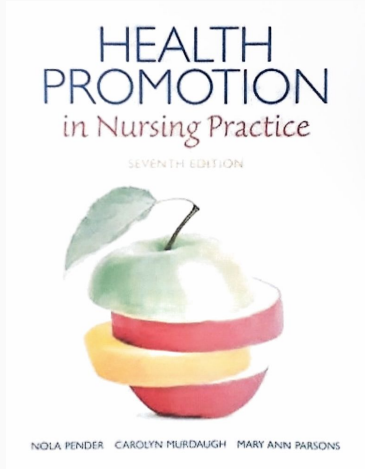
Fan



Zero-DCE

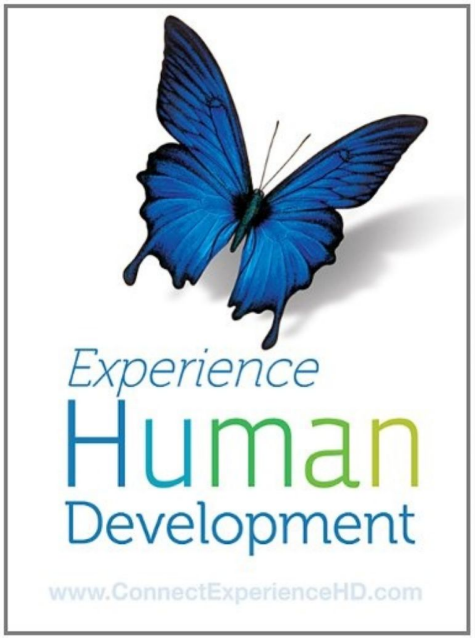


DDocE (Ours)



DDocE (Ours) w/ pyramid

Qualitative Results



PDF

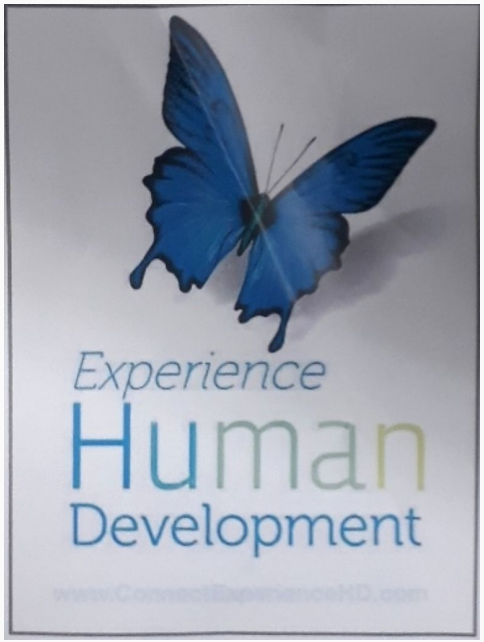
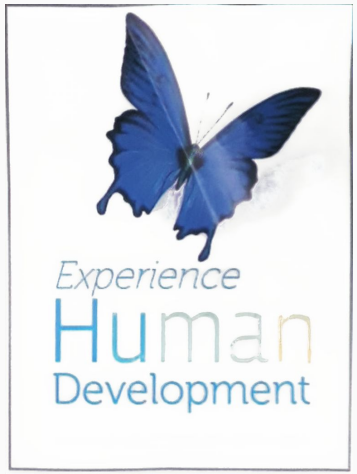


Photo (Input)

Qualitative Results

PDF

Photo Input

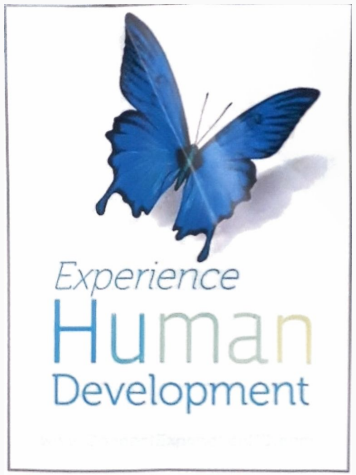


HP Inc.

Fan

Zero-DCE

DDocE (Ours)



DDocE (Ours) w/ pyramid

Ablation Studies

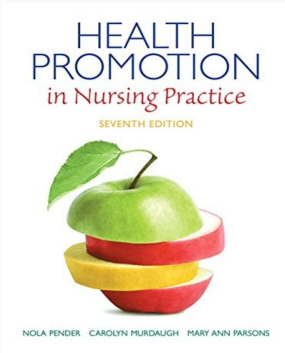
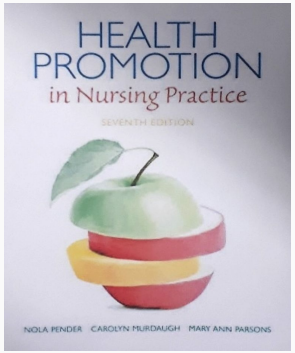
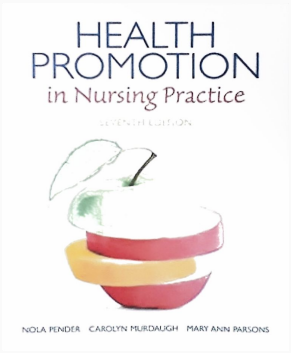
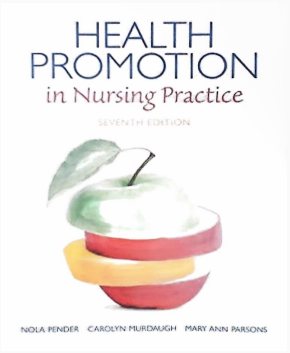
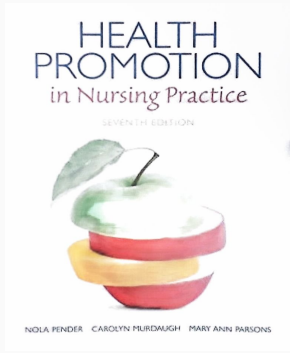
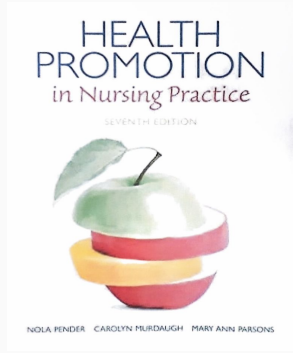
Model	PSNR ↑	MS-SSIM ↑	PieAPP↓	WaDIQaM ↑	LPIPS↓	DISTS↓
DDocE w/ Zero-DCE[1] extractor	15.7 +-3.0	0.81 +-.08	0.94 +-0.8	0.58 +-.07	0.16 +-.05	0.16 +-.04
DDocE w/o squeeze-expand	16.1 +-2.9	0.82 +-.07	**0.92 +-0.8**	0.60 +-.07	0.15 +-.05	0.16 +-.04
Multi-scale aggregation w/ dilation factor of 1	16.1 +-2.8	0.81 +-.08	1.03 +-0.9	0.59 +-.08	0.16 +-.05	0.16 +-.04
DDocE	**16.1 +-2.7**	0.82 +-.07	1.09 +-0.9	0.58 +-.09	0.16 +-.05	0.17 +-.05
DDocE w/ pyramid	16.1 +-3.0	**0.83 +-.07**	1.08 +-0.8	**0.63 +-.06**	**0.14 +-.05**	**0.15 +-.04**

Best result is highlighted in bold.

Documents with colored regions

[1] Guo et al. Zero-Reference Deep Curve Estimation for Low-Light Image Enhancement. In CVPR, 2020.

Ablation Studies

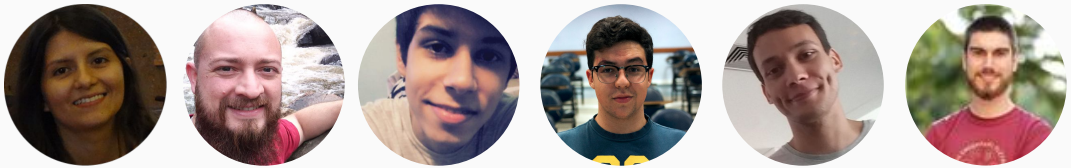
PDF	Photo Input	DDocE w/ Zero-DCE extractor	DDocE w/o squeeze-expand	Multi-scale aggregation w/ dilation factor of 1	DDocE (best result)
					
					

DDocE: Deep Document Enhancement - Summary

- Competitive lightweight method for document enhancement
  - Importance of the components of our model altogether
  - Open challenging scenarios
- Challenges in the evaluation protocol
  - Future work: add human perceptual study

Reach us!

karina.bogdan@eldorado.org.br
guilherme.megeto@eldorado.org.br
rovilson.junior@eldorado.org.br
gustavo.souza@eldorado.org.br
augusto.valente@eldorado.org.br
lucas.nedel.kirsten@hp.com



ELDORADO